

What is claimed is:

- 1 1. A method for improving compression of a stream of data comprising:
2 transforming the data in accordance with a schema; and
3 compressing the transformed data.
- 1 2. The method of claim 1 wherein the transformation step further
2 comprises the step of partitioning the data into a first and second portion which are
3 separately compressed.
- 1 3. ~~The method of claim 1~~ wherein the transformation step further
2 comprises the step of reordering the data into column major order. *INS.* *AI*
- 1 4. The method of claim 3 wherein the transformation step further
2 comprises the step of partitioning the data into columns which are separately compressed.
- 1 5. A method for retrieving a stream of data from a stream of compressed
2 data which has been compressed in accordance with claim 1, the method comprising:
3 decompressing the compressed data; and
4 transforming the data in accordance with a schema.
- 1 6. A method for generating a schema for improving compression of a
2 stream of data comprising:
3 separating a sample of the data into a first portion of low entropy and a
4 second portion of high entropy;
5 partitioning the second portion into columns;

6 searching for combinations of columns that minimize the compressed size
7 of the sample.

1 7. An apparatus for improved compression of a stream of data
2 comprising:

3 means for transforming the data in accordance with a schema; and
4 means for compressing the transformed data.

1 8. The apparatus of claim 7 wherein the transforming means further
2 comprises means for partitioning the data into a first and second portion which are
3 separately compressed.

1 9. ~~The apparatus of claim 7~~ wherein the transforming means further
2 comprises means for reordering the data into column major order.

1 10. The apparatus of claim 9 wherein the transforming means further
2 comprises means for partitioning the data into columns which are separately compressed.